对人类知识和学术的结构性和系统性分析

作者 顾立帆

人类的知识和学术，经历了从经验到科学的发展。其间不断受到蒙昧主义、诡辩等等的困扰和误导。亚里斯多德建立了逻辑学，试图清除诡辩。

二十世纪人类在相对论、量子学、遗传基因研究、和计算机技术方面取得巨大成就。但也发现逻辑学本身有问题，不但无法真正清除诡辩，在计算机和人工智能时代，还会造成更严重的问题。事实上，人类主要的学术分支：数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式的错误。

至少从 leukotomy 开始，脑和智力研究就陷入错误无法纠正的困境。在人工智能兴起的今天，更出现了混淆数学和科学，混淆经验和科学，混淆经验测试和科学实验，把亚里斯多德思维模式当作正确的思维模式等等的错误，在人工智能的哲学基础和测试理论上都存在严重问题 [1]。而拒绝别人用科学实验对经验结论的证伪，也使诡辩又开始在学术界盛行。

本文从分析亚里斯多德思维模式的错误开始，分析 leukotomy 的本质问题和对基因工程的深刻教训，并研究人工智能在哲学基础和测试理论上的问题，澄清了科学和经验及数学的区别，在此基础上用科学的方法对人类知识和学术作结构性和系统性的分析，以此澄清学术上的众多错误，最后导出对语言处理、语言发展、和科学发展的看法。

根据这一系列的研究，可以设计对人工智能的严格控制条件的科学实验，形成智力研究的科学基础，

一、亚里斯多德思维模式在学术研究上的错误

现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重问题和错误。

物理学和化学，本来就是建立在对亚里斯多德思维的批判上。亚里斯多德思维模式凭经验就草率推断。而科学要求严格控制条件的证伪实验作基础。

达尔文的演化理论，说明了物种是动态发展的，亚里斯多德的静态语言方法和思维模式忽视了生命的本质。

亚里斯多德基于特征的生物分类法，是经验性的主观方法。这种主观方法会掩盖分类法中的歧义。现代生物分类法试图引入基因等客观因素，却因此暴露了生物分类法中的歧义是本质性的，无法用逻辑来消除。

基因研究揭示了遗传和变异的复杂本质，但也给生物分类带来进一步危机。基于基因的生物分类虽然比基于特征的生物分类更准确、客观，但基因的渐变、交叉混杂、以及和其他因素的交互作用，也使正确划分物种界线变得困难，甚至在顶层类别划分上都充满争议，其歧义无法用逻辑消除。

幸运的是，生物分类对很多方面的生物研究只具参考价值。很多研究并不真正需要依赖生物分类的精确定义。所以，生物学的很多部分，能够作为严谨的科学。

哥德尔更证明了即使是数学也不符合亚里斯多德的哲学观，其发展需要更高的智力因素。甚至连康德关于人和思维的哲学，都已经突破了亚里斯多德的思维模式。

所以，近年有人试图把亚里斯多德当成学术上的正统，抬高第谷而无视第谷理论体系的错误 [2]，正在把学术引入歧途，特别是在脑和智力研究上。

下面从 leukotomy 开始，分析脑和智力研究上，亚里斯多德思维模式在哲学基础和验证方法上的问题。本文不讨论亚里斯多德思维模式在学术研究以外是否有用。

二、**Leukotomy** 错误的本质及其对基因工程的教训

Leukotomy 切断的脑连接，并不是多余或无足轻重的，而会造成严重问题。问题没有引起足够重视的原因包括：

1）选择手术对象时没有正确区分 personalities 和 mental diseases。

2）给 leukotomy 授与诺贝尔奖时，没有正确地检验 leukotomy 的后果，对提出警告者也没有重视。

3）手术对 personality 和 intelligence 有严重损伤。连大规模实施者 Walter Freeman 都在一定程度上承认。由于缺乏对 personality 和 intelligence 的理解，事后的统计无法反映其损伤 [3]。

无法正确理解 personality 和 intelligence，无法正确区分 personalities 和 mental diseases，无法正确检测对 personality 和 intelligence 的损伤，是由于缺乏正确描述 personality 和 intelligence 的语言，缺乏这种前瞻性的新领域研究所需要的哲学性思考和分析，而造成的无知。

由于目前对 personality 和 intelligence 仍缺乏研究，仍无法正确区分 personalities 和 mental diseases，无法正确检测对 personality 和 intelligence 的损伤，所以对人类的基因改造和基因控制有巨大危险。

Leukotomy 获得诺贝尔奖之前的研究，没有作严格控制条件的科学实验和检测，仅仅凭经验性的手术尝试作出主观的评估。这种主观评估跟后来发现的灾难性后果明显不符。

而缺乏严格控制实验条件的主观评估，也正是 Turing Test 的问题之一。Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，也是主观评估。这个 superhuman 是针对 generic 的 human，既缺乏严格控制条件的实验依据，又不接受别人设计的证伪实验。

Leukotomy 的诺贝尔奖有错不纠，和 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称拒绝严格控制条件的实验证伪，显示 peer review 有严重问题。Peer review 也是主观评估 [4]。

所以，leukotomy 的问题，实际上就是亚里斯多德思维模式的问题。这种问题已经危及脑和智力研究及相关领域。以下第三至五节，讨论跟人工智能有关的各种问题。

三、**AlphaGo Zero** 的 **Superhuman** 宣称不是科学结论

Nature 杂志上的 AlphaGo Zero 论文，宣称 AlphaGo Zero 取得 superhuman 的 performance。但 superhuman 的概念是针对 generic 的 human。该论文并没有提供任何 superhuman 的证据。

AlphaGo Zero 击败 AlphaGo Master 并不是 superhuman 的证据；电脑围棋胜过一些对计算机和人工智能重要的本质问题尚缺乏了解的棋手，也不是 generic 的 superhuman 的证据。这种 game 只是经验测试，不是科学实验。科学实验需要根据科学原理的分析，设计严格控制条件的针对性实验。

科学的基础是证伪，不是经验。比如，不管地球上有多少人看到太阳东升西落，也不能证明太阳回绕地球转。所以不管电脑围棋多少次在 game 中战胜人类，也不能证明电脑围棋有 superhuman 特质。

只要有人设计出能证明 AlphaGo Zero 在围棋上不具 superhuman 特质的证伪实验，就可以否定该宣称。科学结论必须接受别人严格控制条件的科学实验的证伪。但 AlphaGo Zero 至今不接受本人设计的证伪实验。

所以，AlphaGo Zero 论文中的 superhuman 宣称，不是科学结论。这个宣称，破坏了学术语义和科学实验的严谨。这种实验验证的问题，又是亚里斯多德思维模式的问题。

本人无意贬低电脑围棋的进展。相反是鼓励电脑围棋的研究，并将其作为基础智力研究的实验工具。早在 Deepmind 成立之前，我就分析认为电脑围棋可以击败人类顶级棋手，因为围棋是在很小空间里的高度规则化游戏 [5]。这已被证实。但这并不意味电脑围棋有通用智力潜力或超人能力。

AlphaGo Master 击败柯洁后，很多人开始膜拜 AlphaGo Master。而我指出：AlphaGo Master 虽然赢棋，但只要不是最优解，就可能走的是无理棋。这也已被证实，因为 AlphaGo Zero 击败 AlphaGo Master 意味着 AlphaGo Master 很多离谱的走法其实是无理棋。

见到的 AlphaGo Zero 棋谱中，对 AlphaGo Master 的二十局是 AlphaGo Zero 最后版（或稳定版）[6]，有很不寻常现象：AlphaGo Master 执黑的十盘棋中，布局至 16 手完全一样。而 AlphaGo Zero 执黑的十盘棋中，布局至 18 手完全一样，AlphaGo Zero 输棋的 14、16 局，至 45 手完全一样。所以有如下问题：

1）在 Nature Magazime 上发表两篇 AlphaGo 系列文章后，Deepmind 有没有在这二十局棋里把 AlphaGo Zero 和 AlphaGo Master 对局的情况，全面准确地告诉公众？如果答案是，则：

2) 公布的这二十局棋谱中，所出现的布局高度类同，是由于 AlphaGo Zero 接近围棋最优解，还是陷入 local trap？如果认为是前者，需要用严格控制条件的严谨实验来验证。如果是后者，AlphaGo Zero 走的也是无理棋，而 superhuman 的宣称可能对围棋和智力科学研究带来严重误导。

其实 AlphaGo Zero 击败 AlphaGo Master，本身就是人类击败 AlphaGo Master 的证据：人类靠设计新算法系统击败了 AlphaGo Master，而 AlphaGo Master 自身的进化速度远不及人类根据系统结构分析来设计新算法的速度 [7]。Nature 杂志上的 AlphaGo Zero 论文发表前几个月，我就针对 AlphaGo Master 怀疑：利用自我对弈收集数据和 deep-learning 来涨棋，几年后其效果会不会暴露出局限？事实上，AlphaGo Master 的弱点仅几个月就暴露了。

即使人类能设计更新的算法击败 AlphaGo Zero [8]，仍不能证明更新的算法有 superhuman 的特质。我设计的证伪实验也可以证伪 AlphaGo Zero 的后继方案。

我验证 AlphaGo 系列有没有 generic human 的围棋智力的实验并不复杂。围棋下得不好的我也可以做实验，只要 Deepmind 同意在实验上合作。目的是找出人工智能技术普遍而严重的问题。我的实验的更大价值在于较准确地测试 AlphaGo 系列的围棋智力上限，这需要一流人类棋手和 AlphaGo 系列对弈，要有使其暴露弱点的方法。而在正规而公平的围棋比赛中，人类一流棋手是否仍有可能击败 AlphaGo 系列，也需要进一步验证。

我设计的实验，实验结果和结果分析会全部公开。别人可以对实验过程和结果作分析、推敲。这才是真正的科学实验方法。如果人类棋手在我设计的公正实验条件下也能胜过 AlphaGo 系列，意味着 AlphaGo 系列所依赖的人工智能技术有严重的缺陷 [9]。这对脑和智力的研究、以及人工智能的发展都有很大价值。

我设计的实验，基于 Turing Machine 和人工智能的局限性，不需要 AlphaGo 系列论文和棋谱中的信息，所以可以验证人工智能和计算机本身的局限性。但出于对科学和公众负责，论文作者应该公布跟论文有关的全面准确的实验结果：包括 AlphaGo Zero 对 AlphaGo Master，和最后版（或稳定版）的 AlphaGo Zero 对 AlphaGo Zero 的更多棋谱，等等。

基于数据的人工智能或统计方法，无法解析未来的模式演变。如果模式演变不稳定，会产生严重误导。有些以前几乎不出现的问题，模式演变后，甚至可能成为大概率事件。要了解智能的模式演变是否稳定，需要研究智力的基础理论。

四、自动驾驶的定义和测试问题

本人以前写过：“Go gaming is strictly defined within a very small space. Industrial automations are typically designed in well controlled environments, but not strictly defined. Car driving is regulated, but the environment is not well controlled”。

自动驾驶汽车中的问题，远比电脑围棋复杂。如果在电脑围棋中，不稳定的模式演变，都会造成大概率的严重问题，采用类似人工智能技术的自动驾驶就更需要谨慎了。

无人汽车不加限制说明，就是 SAE level 5 自动驾驶，而这是不可能的。这可以通过对智力原理的分析和对人工智能所使用的技术设计实验来验证。而 SAE level 4 的自动驾驶，在不同的驾驶区域，实际上有不同的技术要求。所以 SAE level 4 的定义，会隐含很多问题。即使在同一区域，未来不稳定的模式演变，也可能使已通过 SAE level 4 测试的自动驾驶系统，出现大概率严重问题。

另外，无人驾驶汽车的测试，完全不同于其他汽车。汽车中的动力（燃料或电动）、机械、电子等设计，早就具备坚实的科学基础。而无人驾驶的人工智能，没有科学基础，也未经严格控制条件的科学实验验证，危险性很大。

现在使用的是经验测试，无法解析未来模式的演变。而普通人可能需要很多年跟无人驾驶汽车的大规模互动，才能接触到其临界点，引发重大的模式演变。所以需要对智力的基础理论作前瞻研究，设计严格控制条件的科学实验来验证和证伪这些技术。

本人不直接研究自动驾驶。主张通过对智力的基础研究、和对电脑围棋等等人工智能技术设计证伪实验，为智力研究建立科学基础。能这样在不引发大规模严重后果的情况下，用基础研究的方法，为自动驾驶技术作前瞻性研究，其实是一种幸运。

本人赞成研究电脑围棋，也赞成研究自动驾驶。只是不赞成过度炒作无人汽车，不赞成把未经严谨的证伪科学实验检测的人工智能技术用于商业化的自动驾驶。

下面分析人工神经网络和预测术的本质，分析基于经验和数学的人工智能和科学的本质不同。

五、预测术、人工神经网络、和计算机技术的局限性

人工神经网络实际上是对已有信息的改写和预测。

人类文明的早期，预测术曾被普遍看重。中国古代有用于预测术的易经：《连山》据说是夏朝的易经，《归藏》据说是商朝的易经，以及周朝的《周易》。

商代夏，周代商，模式演变大致是稳定的。旧易经不对，就换新易经。到春秋战国的混乱取代周朝的实际统治时、和秦国结束战国混乱时，模式演变是不稳定的。没有人再能写出流行的新易经。预测术虽仍继续存在，但已远不如夏、商、周时那样重要了。

《道德经》中的“道可道，非常道。名可名，非常名。”实际上是对预测术的否定。即使是现代的科学之道，也需要不断发展。

到欧洲文艺复兴后期，由于天文学的发展，占星预测术又开始流行。

第谷的天文理论体系是错的，却得到了很多资助。开普勒的天文理论体系是对的，却死于贫困。开普勒的天文学研究，在其生前并不受人重视，不得不靠占星预测术来争取收入，维持他的天文学研究。

占星预测术被认为是实用的。而基础天文学研究对一般人没什么用。但天文学能够逐渐发展成为非常严谨的学术，而占星预测术到了今天仍然仅仅是预测术而已。

到二十世纪前期，著名心理学家荣格借助星座分析人的性格。荣格的星座分析，已经和开普勒的占星预测，有了很大的不同。但即使荣格在他的星座分析中引入了很多哲学性的思考，但他的星座学仍达不到哲学完整性，当然更达不到科学的标准。

并不是说预测术就不能研究。但需要区分预测术和科学的区别。也不应该用预测术的研究，来排斥对智力基础理论的科学研究，即使基础科学研究在早期显得没什么用。

预测术是基于经验或数学的。哥白尼、第谷、开普勒时期的天文学，也是基于数学的，还不是严格意义上的科学。开普勒把哥白尼的行星轨道从圆形改成椭圆形，但现代科学发现行星轨道并不是椭圆的 [10]。确定更精确的轨道，需要对力的原理进行分析。所以，哥白尼和开普勒都不是严格意义上的科学家。

爱因斯坦追随的是伽利略的哲学，不是牛顿的哲学。所以科学的创建者是伽利略，并不是哥白尼、开普勒、或牛顿。

科学基于对原理的深入分析，需要用严格控制条件的实验对原理作证伪，并不是对现象层次的预测作验证。所以，科学和经验及数学都有本质不同，也不是预测术。

Universal approximation theorem 的名称，也很容易误导。定理中所能趋近的函数，是有严格限制条件的。并不是所有函数都能无限趋近。而可以抽象成 Turing Machine 及其各种变种的计算机技术，也有本质的局限性。

近年有人根据人工智能的一些成功经验，炒作 Technological Singularity，实际上是对科学哲学的无知。人工智能最大的危险，不是其太强大可能控制或消灭人类，而是大规模应用不成熟的人工智能可能带来巨大灾难。

六、脑和智力研究需要新的科学哲学

理解了科学和经验及数学的本质不同后，就可以分析目前脑和智力研究上的各种问题，澄清哲学和思维模式上的各种错误。

1949 年诺贝尔生理医学奖 leukotomy 存在严重的检测错误，并造成灾难性误导，但至今无法纠正。科学研究上有错不纠，会蔓延成脑和智力研究上的普遍问题。

Turing Test 和主流教科书 AI: A Modern Approach 有哲学基础和测试理论上的严重问题，把 Wind Tunnel approach 应用于人工智能已造成严重后果；权威学术杂志 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，破坏了学术语义和科学实验的严谨；Technological Singularity、In Math We Trust、和"人类正在扮演上帝的角色"的观点[11]，甚至出现在专业研究人士中间；而把脑研究比作登月项目的官方观点，更是惊人的无知 [12]。

种种问题说明：现实中真正奉行的研究思维和方法有严重隐患。科学哲学在 Karl Popper 以后，基本停滞不前，不但导致脑和智力研究在理论上误入歧途，也使科学研究在整体结构上越来越难以突破，甚至错误泛滥。

很多耗资巨大的实验结束后，既证明不了什么本质，也否定不了什么本质，使新的科学理论无法成熟。而对经验性的人工智能技术，拒绝别人的证伪实验，也使诡辩在学术界更加流行。学术语义和科学实验的严谨一旦被打破，教育、学术研究、医疗、和司法等等许多方面，都会出现全局性严重后果，危及未来的发展[13]。

忽视、排斥科学哲学的人，很可能用一种错误的哲学来看待科学，比如经验哲学、实用主义、Occam's Razor、逻辑实证主义，等等。

Occam's Razor 仅仅是一种经验的哲学方法，不但不是科学，也不是严谨的科学哲学。所以从几种无法证伪的理论中，选择最简单的理论，并不是科学结论。对 Occam's Razor 在学术研究中的广泛使用，需要科学哲学的新研究来澄清其隐含限制、限定其使用。

科学理论的简单和复杂，甚至不一定存在数学上的全序关系。所以，几种暂时无法证伪的科学理论中，也不一定存在最简单的理论。即使存在数学上的全序关系，Occam's Razor 所剃掉的是不是真正多余的部分，有可能需要很长时间才能发现，不能草率下结论。像 Leukotomy 切断的脑连接，并不是多余或无足轻重的。就是一个深刻教训。

科学的基础是证伪，不是经验。不但亚里斯多德思维模式在学术研究上是错误的，连逻辑实证主义也有问题。逻辑判断使用的语言和经验证据之间可能有很大的差别，经验证据往往不能真正支持逻辑判断。而科学的原理分析，可以发现两者之间的差别。基于科学原理分析的证伪实验，可以检验这种区别。

所以，对人类在历史长河中形成的林林总总的经验、知识、学说，及研究领域中的很多流行观点等等，需要从科学哲学的新高度去分析、理解、判断其对错和长远价值。对人工智能的实用性、可靠性、安全性，也需要用严格控制条件的科学实验来证伪。

以下两节介绍本人对人类知识和学术的结构性和系统性的分析，作为智力研究的科学哲学基础。而把智力研究提升到科学层次，则需要严谨的科学实验基础。Gu Test: A Progressive Measurement Of Generic Artificial Intelligence 一文介绍的是本人的实验

设想。

七、对人类知识和学术的结构性分析

语言是人类重要的智力特征。其他已知生物，都不具有复杂的语言能力。

在语言基础上，人类发展出两类重要的学术：数学和哲学。数学向精确化发展，哲学向完整性发展。

科学起源于哲学。所以，严格意义上的哲学和科学，具有哲学完整性。科学也不断引入数学和精确性，但科学的本质特点是严谨的证伪实验基础。数学达不到科学标准，也不具有哲学完整性。

达尔文的演化理论，可以通过科学实验作验证，所以是科学理论。而他的物种起源理论和宇宙大爆炸理论虽从科学理论推导而来，却仅仅是历史哲学层次的准科学。虽然会用到很多用科学方法收集的证据，但无法用严格意义上的严格控制条件的科学实验来证伪。所以达不到科学理论的标准。

生物分类法的本质歧义，无法用逻辑来消除。所以达不到哲学完整性，仅仅是语言层次的知识，只能作粗糙的参考。最多算准哲学。很多重要的生物学理论并不依赖生物分类的精确定义，所以可以作为严谨的科学。

研究智力基础理论，需要精确理解和区分语言、哲学、数学、科学等等不同层次和标准的理论和方法，研究其各自发展的基本原理。

现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重问题和错误。现在需要跨学科对亚里斯多德思维模式的问题和错误，作科学哲学上的剖析和总结，作为智力研究的哲学基础。

开普勒的占星预测术、荣格的星座分析学、人工神经网络，也达不到哲学完整性，当然更不是科学。

对理论标准，作语言层次、哲学层次、数学层次、科学层次上的严格细分，是为了建立严谨的科学哲学和科学体系，和精确地衡量人类智力的不同范畴和方法、及所作结论的不同性质。

下面对数学和科学的历史演变及其重要临界点作分析，作为对人类主要学术体系的系统性理解基础。

八、数学和科学的演变及其重要临界点

从毕达哥拉斯到伽利略，是人类文明走上可持续发展的关键。

毕达哥拉斯的抽象化思维，创立了数学。他对算术和数的定义和体系，按现代抽象代数的概念也是成立的，但不适用于几何和其他一些系统。毕达哥拉斯算术系统本身在科学上是错的，但在数学上又是可以成立的、有价值的、可以继续发展的。所以数学不同于科学，也不能简单用数学来验证科学上的对错。

希帕索斯的发现，本来也是对数学和毕达哥拉斯体系的发展。但希帕索斯事件之后，毕达哥拉斯学派本身就没有太多成就了。其他群体则能继续他们的工作，后来居上。

欧几里德几何创建了重要而又可以完备化的数学理论体系，并带动了阿基米德等等一大批泛希腊化时代的学术研究者。

毕达哥拉斯和欧几里德两千年之后，有牛顿和莱布尼茨创建微积分，罗巴切夫斯基等人推翻欧几里德几何的唯一性，建立非欧几何 [14]。

朗道实际上把自己当作第四流的科学家：牛顿是第一流，爱因斯坦是第二流，创建量子物理的才俊们是第三流，和朗道同时代的美国物理学家理查德·費曼也是第三流。也许朗道是自谦，但他肯定无法和爱因斯坦相比。而俄罗斯历史上是出过门捷列夫的。

不知朗道该如何评价伽利略、达尔文。

按照我对学术成就的评价，创建重要学科、和重要学科中的理论体系，是第一流的工作；在重要学科中提出革命性理论体系，是第二流的工作；对主要理论体系的重要延伸或修补、或开创较小的学术分支，可列入第三流的成就。第一、第二流的学术成就非常罕见，第三流的工作已经非常了不起了。这个标准跟朗道的评价高度吻合。

物理学中，伽利略和牛顿的贡献，跟数学中的毕达哥拉斯和欧几里德相仿，属第一流。相对论和量子力学是重要学科的革命性理论体系，跟数学中的微积分和非欧几何相仿。所以，爱因斯坦的贡献是第二流的，而创建量子力学有许多人的贡献，他们的贡献只能算第三流的。量子电动力学只是小分支，所以理查德·費曼的工作也是第三流的。

牛顿和莱布尼茨在数学中的贡献是第二流的。但非欧几何的创建，由多人逐步完成，有许多人的贡献。罗巴切夫斯基和鲍耶对推翻欧几里德几何的唯一性最重要，他们和希帕索斯（发现无理数）、亚里斯多德（逻辑学）、帕斯卡（概率论）、哥德尔（哥德尔定理）相仿，仅略逊于牛顿和莱布尼茨。

拉格日朗（数学分析的成熟化）、图灵（图灵机作为计算模型）的贡献是第三流的；数学中的高斯，和物理学中的朗道，能不能挤进第三流，我不知道；

祖冲之的贡献，未必能超过高斯。但有观点认为《九章算术》首次引入负数 [15]。此说如果确实，《九章算术》的数学成就，也许并不比希帕索斯和亚里斯多德逊色很多。这可能跟中国古代的阴阳玄学哲学观念有关 [16]。

数学是非常古老的学科，现在看起来很重要的一些成就，跟数学历史长河中的很多大师比起来，不一定更重要。十七、十八世纪的数学，和二十世纪前半期的物理学类似，已经十分兴旺，但基础是更早的人奠定的。

达尔文跟毕达哥拉斯和伽利略相仿。孟德尔不如欧几里德和牛顿，因为生物学整体理论在不久之后还需要几次重要的综合才能形成。

波义耳和门捷列夫在化学中的作用，有点类似于物理学中的伽利略和牛顿。但门捷列夫和波义耳不但时间相隔较久，而且中间有很多人的重要贡献。所以，他们的学术贡献，也不如伽利略和牛顿。

这里的评价，全部按照学术贡献的大小。不是评估个人智力，而是评估对人类整体智力发展的影响。人各有特长，个人智力也不一定都表现出来，很多情况下是不可比较的。而学术贡献，是已完成的工作，可以客观分析。

整理历史上学术贡献的大小，一方面是为了研究人类智力的演变和其中重要的临界点，另一方面也是研究能促进未来发展的各种因素。现在把智力研究转化成科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。

如果物理学、化学、生物学、人工智能或现有数学等等，都不能提供这些概念、基础、和形式，就需要创建新的学科。能不能提供，可以通过证伪实验来验证。

科学和数学的本质和方法，有所不同。科学兴起虽然较迟，但能弥补数学思维方法的一些严重不足。事实上，物理科学比数学更依赖于哲学性思考。而生命科学的哲学性，比物理科学更重。开创智力新科学，比起生命科学也更具哲学上的挑战性。

从数学、到物理科学、到生命科学、到智力科学，随着哲学性思维比重的不断加大，其研究和教育也会越来越平民化，可以为普通人带来更多的机会。因此也需要探索智力科学时代平民化教育和研究的新方式。

新的哲学思维，也会带来思想上的新挑战。而希帕索斯和伽利略因为追求学术真相而受迫害，更应引起人们的警觉。相比之下，高斯虽是优秀的数学家，却缺乏追求学术真相来推动进步的求道精神，历史地位有限。

公正的研究环境、和免受学术迫害，是科学能够持续发展的前提。

九、语言处理、语言发展、和科学发展

语言是人类知识的重要载体，也是哲学、数学、和科学的基础。

根据对人类知识和学术的结构性和系统性分析，可以建立对语言处理、语言发展、和科学发展的新理论。

人类知识和学术的发展，依赖于语言的发展，特别是在重要的临界点上。

人工神经网络本质上是对已有信息的改写和预测，在低层次的自动翻译上可以取得很多进展。但这种预测并不是科学方法，无法发展科学的新思想，也无法提供科学发展所依赖的语言发展。

人工智能和计算机技术，有 Chinese room 的问题，也无法正确处理高阶逻辑、识别诡辩和亚里斯多德思维模式等等的问题。

人类和人工智能、计算机有本质的不同。其各自的长处、短处也很不相同。所以，用给人类考试的试题来测试人工智能，参考价值很有限。实际上反而会掩盖人工智能的很多关键性问题，很容易引起误解。

现在测试语言理解的 dataset 如 SQuAD、CoQA、QuAC、和 GLUE 等等系列，只能检测较简单的阅读理解，不能测试用语言来深入思考的能力。因为缺乏对语言智力的科学和结构性分析，这些 dataset 无法真正衡量人工智能和人类智力的差距，也不能在诡辩识别、高阶逻辑处理、Chinese room 等问题上对人工智能有多少提高。

NLVR^2（Natural Language for Visual Reasoning for Real）等 testset 虽然难度较高，但仍缺失对重要语言智力因素的测试。GLUE 也并不具有测试 generic 语言智力的能力。

事实上，所有这些 Testsets 都陷入亚里斯多德思维模式的陷阱，都不是严格意义上的科学测试方法。所以说，主流教科书《AI: A Modern Approach》在哲学基础和测试理论上的错误，覆盖了人工智能整个行业。

现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重错误。如果不能识别诡辩、亚里斯多德思维模式等等的问题，而错误地把人工智能用作 guideline 或 judgement，会误导司法、医疗、研究、和教育等等，损害智力的发展。

所以需要研究语言、哲学、数学、科学的发展，研究智力的基础理论，建立对语言处理的科学测试方法，验证 Chinese room 问题，检测对高阶逻辑的正确处理、和识别诡辩及亚里斯多德思维模式等等问题的能力。

十、结语

科学哲学在 Karl Popper 以后，基本停滞不前。导致学术界错误泛滥，也使脑和智力研究在哲学基础及测试理论和方法上都误入歧途。

本文引入对人类知识和学术的结构性和系统性分析，作为脑和智力研究的科学哲学新基础。以此分析学术上的种种错误，设计严谨的科学实验对人工智能技术证伪。

把智力研究转化成科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。科学的生命力在于设计可验证的理论。经验测试不是科学实验，科学结论必须接受别人严格控制条件的科学实验的证伪。

对智力的基础科学研究，不但可以帮助人类和人工智能更好地配合，而且可以带动脑科学、生物学、语言学、心理学、哲学、数学等等一大批学科的发展，为多样化的教育带来新的机会。

学术界流行的很多观点，破坏了学术语义和科学实验的严谨。如果不纠正，会使诡辩泛滥，在教育、学术研究、医疗、司法等等许多方面，引发全局性的危机。

在 The Development of Liberal Arts and Sciences 一文中，本人将分析蒙昧主义和诡辩的土壤，以及科学对人文的依赖。

～～～～～～～～～～～～～～～～～

[1] 人工智能的主流教科书 AI: A Modern Approach 的第三版，从哲学基础中删掉了对苏格拉底和柏拉图的介绍，把亚里斯多德思维模式

当作正确的思维模式：“Aristotle... was the first to formulate a precise set of laws governing the rational part of the mind.”(On page 5)。

由于这个思维模式的错误，该书提倡的 Wind Tunnel approach 也只是经验方法，不能科学地测试人工智能技术的可靠性和安全性。

该教科书在哲学基础和测试理论上的错误，实际上也覆盖了整个人工智能行业。

[2] 事实上太阳系的重心就在太阳附近。以太阳为基点的参照系比以地球为基点的参照系更接近惯性参考系。如果采用地心说或第谷的理论，就无法发展出牛顿的物理体系，无法渐进地发现科学规律。

第谷的观察工作仍有价值是侥幸，因为地心说、日心说、第谷理论在观察方法上没有争议。哥白尼、第谷、开普勒时期的天文学，还不是严格意义上的科学，所使用的方法也不是严格意义上的科学方法。

而很多科学实验，需要严格控制实验条件。实验结论要根据实验条件来推导。错误的实验条件会导致实验无效或错误的结论。

[3] 按政府报告 Leucotomy in England and Wales, 1942-1954，该手术的 9284 名患者中，41%恢复或有明显改善，28%稍微改善，25%显示无变化，2%恶化，4%死亡。

但这个统计很容易被误解。如此破坏性的脑手术，当然会破坏 personality 和 intelligence。所以，这里的 25%显示无变化，不是指 personality 和 intelligence 未受破坏，仅指为控制精神状态而特殊定义的 clinical condition 无变化。同理，41%恢复或有明显改善和 28%稍微改善，也只是特殊定义的 clinical condition 有改善，personality 和 intelligence 实际上都有损伤。

本来，leucotomy 手术就是为了控制精神状态，而不惜损伤脑功能。

也有冠名 Renato M.E. Sabbatini 教授的文章指：Even lobotomy's preponents admitted that only one third of the operated patients would improve, while one-third remained the same, and one-third got worst 。这些数据来源不明，跟 Leucotomy in England and Wales, 1942-1954 的差距也很大。（http://www.cerebromente.org.br/n02/historia/lobotomy.htm）

但这里的 one third would improve 和 one-third remained the same，也仅仅指特殊定义下的 clinical condition，不是指 personality 和 intelligence 未受破坏。

所以，由于缺乏正确描述 personality 和 intelligence 的语言，就无法对 leucotomy 手术和使用基因技术的 BRAIN Initiative 方法的破坏性后果作研究。这种情况急需改变。

[4] 对学术论文的审核，只代表审核者的观点，不是学术上的定论。所以，学术上的苏格拉底方法很重要。要给被审核者和其他人反驳的机会。

对科学论文的审核，也跟数学论文不同。数学论文只能靠 peer review。科学论文中的结论，可以用新的证伪实验来推翻。其错误，从论文本身可能看不出。所以，对科学论文的 peer review，并不能保证其正确性。要让别人有机会设计新实验来证伪。

对科学论文的审核，可以发现推理问题，实验数据是否真实等等。但推理问题只能引起注意，有可能是出于理解有误。所以下结论要经过对话、反驳、和实验证伪。AlphaGo Zero 论文中的 superhuman 宣称，是针对 generic 的 human，并没有相应的证据。本人却可以设计严格控制条件的实验来证伪此宣称。但 AlphaGo Zero 方面至今拒绝对话和实验。

[5] 许峰雄在 Cracking Go 文中更早提出。但他的依据很多来自开发 Deep Blue 的经验，对围棋不一定成立。我的理论分析，不依赖经验，反而能成立。我的理论分析还可以进一步延伸：AlphaGo 系列虽然赢棋，但只要不是最优解，就可能走的是无理棋；可以设计严格控制条件的严谨实验来检验 AlphaGo 系列是否有通用智力潜力，是否超人。

[6] 根据 http://www.alphago-games.com/，AlphaGo Zero 对 AlphaGo Zero 中只有一局是最后版（或稳

定版）的。在 https://www.101weiqi.com/chessbook/player/38348/也有相关棋谱。

[7] AlphaGo Master 不是生命，AlphaGo Master 进化快慢，也跟生物的演化本质无关。部分原因是裁定围棋胜负的规则是由人类定义的，而生物演化的规则并非由人类定义。

[8] http://www.alphago-games.com/上，Full Strength of Alphago Zero, i.e. Final Form，被称为 40 Blocks。20 Blocks 被称为 Not Full Strength of Alphago Zero。有更新结构的系统能击败 Alphago Zero 也是非常可能的。

[9] 我选择围棋作实验工具，是因为可以用围棋验证人工智能和计算机的一些基础科学理论，跟军事战略和战争艺术无关。围棋是和平的竞技，高度抽象而又直观。既古老又现代，能吸引亚洲以外的电脑研究。已经成为跨文化研究、交流的平台。

第一轮实验后，开发者可能修改软件。所以有些实验证据只能在第一轮实验中收集。第一轮实验需要跟最强版本电脑围棋作较大规模的实验，以收集足够的实验证据。

我对 AlphaGo 系列的实验不是白盒测试，不涉及知识产权。如果 Google 愿意公正地检验，可以仿照美国国会图书馆先例，先向国会图书馆递交 AlphaGo 柯洁版、AlphaGo Zero、和最新 AlphaGo 版本所需要的所有硬件、软件和文档，能捐款给图书馆用于管理和维护则更好。以后在第三方场地实验时，可以向国会图书馆核对软件、硬件和文档等版本。Human level artificial intelligence 的检验，是人类发展史上的里程碑。如果无人能对 AlphaGo 系列证伪，国会图书馆就有继续保留其版本的价值。其他人有实验想法，可以先做。如果其他人都失败了，就没有道理拒绝我的实验方案。

这也为美国政府了解人工智能的最新进展提供了渠道，并为以后的人工智能测试建立了一个范式。一旦相关实验设施建立好并熟悉了这种流程，以后重要的人工智能测试费用会大大降低。

[10] 由三体问题的研究发现，行星的真正轨道极为复杂，传统数学都难以完整表达或逼近。所以，下文提到：Universal approximation theorem 的名称，也很容易误导；而可以抽象成 Turing Machine 及其各种变种的计算机技术，也有本质的局限性。

[11] In God We Trust 是美国的格言。这里提到的 God 是全能神。现代科学可以构造多个具有哲学完整性且无法证伪的科学理论。这些无法证伪的科学理论，使全能神的存在，也无法被证伪。但人不是全能神，人无法完整理解全能神，所以说 Trust。

数学是人的一种智力理论，不是全能神。数学的局限性，早已被证明。所以，In Math We Trust 并不正确，是迷信。In Math We Trust 在精英专业研究人士中的流传，暴露了科学哲学教育的缺乏和现代教育的问题。

[12] 登月项目是工程。需要定义明确的实施目标。而脑研究是基础科学研究，需要通过科学实验的证伪来不断修改假设。所以，工程和基础科学研究的方法完全不同。出现这样的混淆，是教育和研究严重失败所造成的无知。

[13] 在政治宣传、商业宣传中，诡辩和语义的混淆比比皆是，难以清除。所以，科学研究和司法，需要跟政治宣传和商业宣传严格区分，保持应有的独立性。

[14] 高斯不但没有公布他自己的非欧几何想法，也没有公布 1819 年就交给他的 Ferdinand Schweikart 的类似想法。所以高斯对非欧几何，不但无功，反而有过。原因就在于高斯太注重坚实的技巧性，不了解尚未成熟的思想体系的价值，也不重视苏格拉底方法对学术发展的作用。高斯不但缺乏学术眼光，也达不到罗巴切夫斯基的思想境界和层次。

罗巴切夫斯基和鲍耶推翻欧几里德几何的唯一性，已是 1830 年前后了；他们在非欧几何上贡献最大。黎曼几何在物理相对论研究中，有价值。但黎曼在非欧几何中，只是作延伸性工作，贡献次于罗巴切夫斯基和鲍耶。贝尔特拉米为非欧几何带来新的语义和完整性，也很重要。

由于被高斯埋没，Ferdinand Schweikart 可能没有对非欧几何的发展产生多少实际影响。我也不知道如何评价他的工作。

[15] 《九章算术·方程》：“正负术曰：同名相除，异名相益，正无入负之，负无入正之。其异名相除，同名相益，正无入正之，负无入负之。”

也有观点认为，《九章算术》之前，中国已经出现负数概念。

[16] 中国古代百家争鸣时期，就有阴阳家，《道德经》和解释《易经》的《十翼》也都论及阴阳。阴阳概念的真正起源，有待考证。但《易经》体系的阴阳并不导出负数，而是：老阳数九，老阴数六，等等。

阴阳概念，不但可以很容易导出正负数，也为现代正反物质的理论提供了一个很自然的通俗理解语言和思维模式。但能相互抵消的正负数，是在阴阳概念基础上的飞跃。现代正反量子湮没的理论，又是对可抵消的正负数概念的飞跃，因为量子湮没是伴随有能量形态的转化的。从中正好可以看到从哲学、到数学、到科学的思维模式的演变和发展。

但阴阳不是科学哲学语言，进一步发展的潜力有限。在形成正负数以后，中国古代数学也没有在这个方向有多少后继发展。所以，即使哲学能帮助形成一些新的思想，哲学本身也需要不断发展，需要新的语言概念。“道可道，非常道。名可名，非常名。”，至少说明不能固步自封。拒绝别人的质疑和证伪实验，也是在拒绝进一步发展的机会。

中国古代没有像希帕索斯那样对无理数的准确发现，对类似无理数的概念没多少抵触，但也不理解其价值。这里有哲学和思维模式的不同，受此影响智力的发展轨迹也不同。